Salient object detection via reciprocal function filter

Chen, Wenbing; Dahal, Keshav; Huang, Shuxian

Published in:
IET Image Processing

DOI:
10.1049/iet-ipr.2018.5722

E-pub ahead of print: 18/04/2019

Document Version
Peer reviewed version

Salient Object Detection via Reciprocal Function Filter

Wenbing Chen[1*], Keshav Dahal[2], Shuxian Huang [1]

[1] College of Math and Statistics, Nanjing University of Information Science and Technology, 210044, Nanjing, China
[2] School of Computing, Engineering and Physical Sciences, University of the West of Scotland, PA1 2BE, Paisley, UK
[*]chenwb@nuist.edu.cn

Abstract: Salient object detection (SOD) has been attracting a lot of interest, and recently many computational models have been developed. In this paper, we formulate a SOD model, in which saliency map is computed as a combination of the colour, its distribution-based saliency and orientation saliency. Similar to traditional SODs, the proposed method is based on super-pixel segmentation and super-pixelutilizes both colour and its distribution-based saliency to generate a coarse saliency map. However, distinct from traditional SODs, we further use orientation contrast to optimize the coarse saliency map to obtain an improved saliency map. Our contributions are twofold. First, we combine colour uniqueness and its distribution with local orientation information (LOI) used in Itti's model to effectively improve profiles of salient regions. Second, a reciprocal function is defined to substitute the Gabor function used in LOI and we have proved that the substitution could detect relatively homogeneous and uniform regions at the boundary of salient object, whereas it is what the traditional models lack. Our approach significantly outperforms state-of-the-art methods on four benchmark datasets while, we demonstrate that the proposed method runs as fast as most existing algorithms.

1. Introduction

Salient object detection aims at accurately and uniformly detecting objects that grab human attention in images. Salient object detection methods commonly serve as the first step for a variety of computer vision applications, including object segmentation [1], image compression [2], object recognition [3], image retrieval [4], etc.

Traditional saliency detection methods rely on various saliency cues. The most widely explored one is contrast, which has been shown to be the most influential factor to visual attention in the human vision system [5], [6]. Local and global contrast has been successfully adopted to derive saliency maps in various saliency detection methods, where the definition of contrast is based on various types of handcrafted image features (e.g., colour, intensity and histogram) at pixel or super-pixel scale [7], [8], [9].

Some recent works also utilize various prior knowledge as informative saliency cues. Background prior [9-11] hypothesizes that regions near image boundaries are probably backgrounds. However, it often fails when salient objects touch image boundaries or have similar appearance with backgrounds. Compactness prior [12] advocates that salient object regions are compact and perceptually homogeneous elements. Objectness prior [13, 14] tends to highlight an image region which is likely to contain an object of a certain class. Although these priors can further provide informative information for salient object detection, they are usually explored empirically and have to be carefully tailored for adapting to different types of image data with a wide variety of objects and their contextual interactions, thereby making them less applicable to a wide range of problems in practice.

In recent years, deep convolutional neural networks (CNNs) [15] have been employed in salient object detection to obtain more robust features, and have achieved substantially better results than traditional methods [16], [17]. Features extracted using CNNs contain more high-level semantic information since those CNNs were typically pre-trained on datasets for visual recognition tasks. Generally, these methods can achieve favourable performance when the ground truth annotations of the training samples are given, but are time-consuming at the cost of high computational complexity and heavily rely on the properties of training data.

In this paper, we formulate a salient object detection model, in which saliency map is computed as a combination of the colour, its distribution-based saliency and orientation saliency. Our approach consists of six steps. The first one is pre-segmentation, which segments an input image into multiple regions. Second, for each region we calculate its colour contrast and saliency, which illustrate that a salient object should have strong contrast to their surroundings. Third, we define second saliency, which is called colour distribution, to enhance the saliency of super-pixels belonging to the salient object and suppress the saliency of super-pixels belonging to the background. Fourth, we present a novel reciprocal function filter to generate orientation saliency. Fifth, we integrate uniqueness, distribution and orientation contrast into the final saliency map. Finally, for the fused saliency map: a) super-pixelwe loop the first step to the fifth step to generate N scale saliency maps $\{S^{(1)},S^{(2)},\cdots,S^{(N)}\}$ based on each of N different scale super-pixel segmentation of the input image; b) saliency map fusion algorithm in Sec. 3.6. is applied to fuse N scale saliency maps and yield the fused saliency map.

In summary, this paper has the following contributions:

• We combine colour uniqueness and distribution with local orientation information (LOI) used in Itti's model [18] to effectively improve profiles of salient regions. The Itti's model is discussed in the related work section.

• We define a reciprocal function to substitute the Gabor function used in LOI, and have proved that the reciprocal function could detect relatively homogeneous and uniform salient regions and achieve a remarkable improvement for the saliency detection.

• We develop a complete saliency framework by integrating our saliency model with multiscale image segmentations.

The rest of this paper is organized as follows. Sec. 2 introduces related work and discusses their differences with our proposed method. The saliency computation framework is presented in Sec. 3. Sec. 4 shows experimental results to substantiate the effectiveness of the proposed method. Finally, conclusions are drawn in Sec. 5.

2. Related work

According to surveys [19, 20] presented by Borji et.al., saliency detection methods are categorised into bottom-up and top-down. Since the proposed saliency detection method used in our framework belongs to the former, here we only review related bottom-up methods. For top-down category, we suggest to refer to surveys [19, 20].

As an earlier and original creative work, Itti et.al. [18] proposed a centre-surround model, which yields saliency map by using three local feature contrasts for intensity, colour and local orientation information of an image. The three centre-surround differences are generated by like-DOG (Difference of Gaussian). So far, the model still is a prototype for the salient object detection. Specially, since 2010, saliency detection has been made many significant progresses and many novel methods have been continuously proposed. According to Borji et.al.'s surveys, in existing methods, region-based salient object detection methods are increasingly popular with the development of super-pixel algorithms. For these methods, their general frameworks are that an input image is first over-segmented into many super-pixels, and then based on these super-pixels regional saliency maps are derived. These regional saliency maps are defined as uniqueness in terms of global regional contrast and widely studied in these existing methods. Cheng et.al. [7] proposed a regional-contrast-based saliency extraction algorithm, which simultaneously evaluates global contrast differences and spatial coherence. However, saliency maps obtained using their methods may contain background clutter and sometimes highlight parts of the object. Perazzi et.al. [12] combine colour contrast and colour distribution to perform saliency detection. They show that complete contrast and saliency estimation can be formulated in a unified way using high dimensional Gaussian filters. Then, an up-sampling procedure is carried out to assign saliency value to each pixel. However, there exists an issue similar to Cheng's model, i.e., sometimes highlighting parts of the object. Yan et.al. [21] proposed a multi-layer approach to analyse saliency cues, which mainly tackles an issue that detection accuracy could be adversely affected if salient foreground or background in an image contains small-scale high-contrast patterns. Jiang et.al. [22] presented saliency detection via absorbing Markov chain on an image graph model, which jointly considers the appearance divergence and spatial distribution of salient objects and the background. However, the method preserves object boundary not well. Li et.al. [23] proposed a visual saliency detection algorithm from the perspective of reconstruction errors, in which image boundaries are first extracted via super-pixels as likely cues for background templates, from them dense and sparse appearance models are constructed. Zhu et.al. [24] proposed saliency optimization from robust background detection, which integrates the boundary prior or background information with other cues to yield saliency map.

Inspired by the feature integration theory, some approaches focus on learning the linear fusion weight of saliency features. Liu et al. [25] propose to learn the linear fusion weight of saliency features in a Conditional Random Field (CRF) framework. Recently, the large-margin framework was adopted to learn the weights in [26]. Due to the highly non-linear essence of the saliency mechanism, the linear mapping might not perfectly capture the characteristics of saliency. In [27], a mixture of linear Support Vector Machines (SVM) is adopted to partition the feature space into a set of sub-regions that were linearly separable using a divide-and-conquer strategy. Alternatively, a Boosted Decision Tree (BDT) is learned to get an initial saliency map, which will be further refined using a high dimensional colour transform [28]. In [29], generic regional properties are investigated for salient object detection. Li et al. [30] propose to generate a saliency map by adaptively averaging the object proposals [31] with their foreground probabilities that are learned based on eye fixations features using the Random Forest regressor. Wang et al. [32] learn a Random Forest to directly localize the salient object on thumbnail images. Moosmann et al. [33] utilize a saliency map to guide the sampling of sliding windows for object category recognition, which is online learned during the classification process.

In recent years, with neural network and deep learning continuously made progresses, many deep neural network models for salient object detection have been proposed. Huang et al. [35] formulate saliency detection problem as a multiple instance learning (MIL) task, where the object proposals and super-pixels are taken as bags and instances respectively. Jiang et al. [36] propose a supervised learning approach which utilizes the structural SVM framework and formulates the salient object detection and existence problems jointly in a single integrated objective function. He et al. [34] utilize a novel superpixel wise convolutional neural network approach, which called SuperCNN to learn the internal representations of saliency in an efficient manner. Li et al. [37] propose a multi-task deep saliency model based on a fully convolutional neural network with global input and global output. Li et al. [38] propose an end-to-end deep contrast network which consists of two complementary components, a pixel-level fully convolutional stream and a segment-wise spatial pooling stream. Liu and Han [39] propose an end-to-end deep hierarchical saliency network based on convolutional neural networks, which learns various global structured saliency cues and then hierarchical recurrent convolutional neural network (HRCNN) is adopted to further hierarchically and progressively refine the details of saliency maps. Hou et al. [47] propose a new method that provides rich multi-scale feature maps by introducing short connections to the skip-layer structures within the Holistically-Nested Edge Detector (HED). However, even though for state-of-the-art deep learning models for SOD similar to [37, 38, 39], there

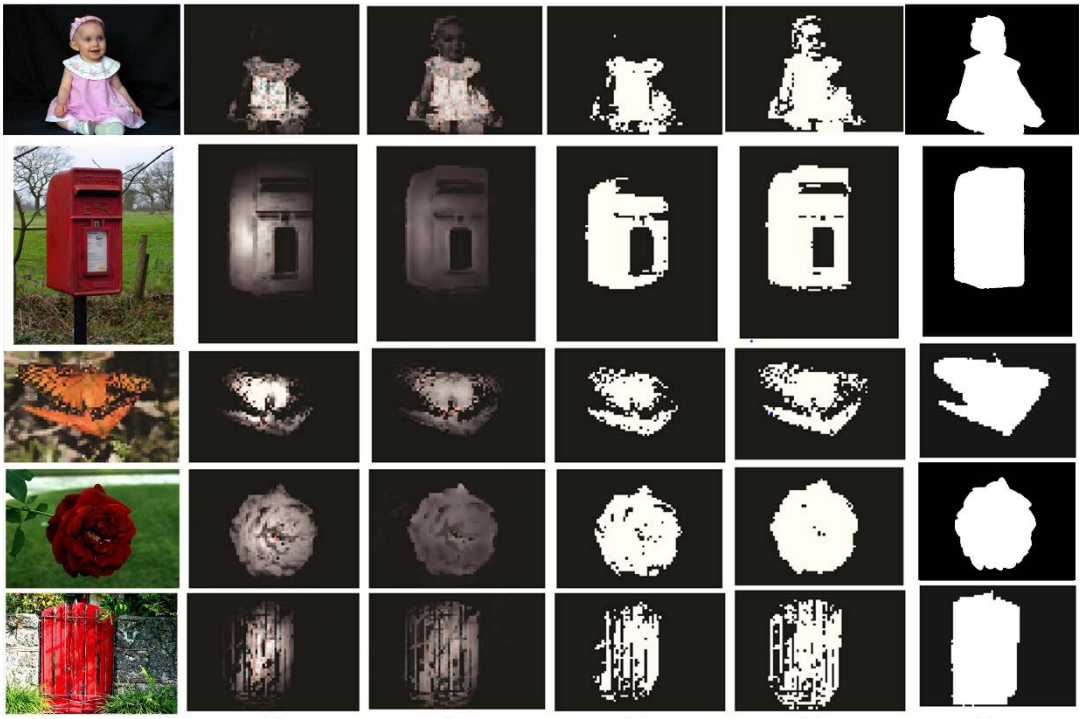

Fig. 1. *Comparison between salient filter (SF) and SF with reciprocal function (SFR): a) source images; b) saliency maps yielded by SF; c) saliency maps yielded by SFR; d) threshold results yielded by SF; e) threshold results yielded by SFR; f) the gound truth. Where d) and e) are yielded by setting threshold* $\boldsymbol{T} = \frac{2\times\sum_{i=1}^{W}\sum_{j=1}^{H}S(i,j)}{W\times H}$ *for their corresponding saliency maps, W and H are width and height of each saliency map, S(i; j) is the pixel at the point (i, j) in each saliency map.*

exists three deadly drawbacks: 1) a large scale training dataset with labels must be accumulated in advance; 2) for each label, i.e., ground truth, in the training dataset, essentially it must depend on handcrafted SOD method to produce; 3) a deep learning model can be used to detect salient object from the input image only when it has been trained, however, it takes long time for training deep neural network.

Therefore, handcrafted SOD method is still a fundamental research. The proposed saliency detection method is a handcrafted SOD method, which is motivated by the two models in [12] and [18], but it obviously differentiates from them and focuses on how to produce high quality saliency maps to totally highlight the entire object and ensure the boundary of the detected salient object has a better fitness to its corresponding ground truth.

3. The saliency detection model

Perazzi et. al. [12] proposed a saliency detection method, which has attracted many researchers to pay attention due to its simple and clear idea. The model simply uses two cues, colour and its distribution to detect salient regions and yield saliency map for a given image. It has a higher efficiency and a lower computational complexity. However, as shown in Figure 1 (b) and (d), many experiments have frequently demonstrated that some detected salient regions lack of integrity, especially near the boundary of salient region, not enough smooth, even some parts near the boundary of salient region are lost, see the little girl at the top row in Figure 1 (b) and (c), her head part has not been able to be detected in the corresponding saliency map and binary image. As it can be seen in Figure 1, other images also expose a similar phenomenon. Through investigating a reason why the phenomenon happens, we found that it is due to the so-called bottom-up model. As mentioned above, for any bottom-up saliency detection model, it walks through such a process: 1) it segments an input image into many small regions or super-pixels, in fact the input image is transformed into a coarse-grain image since all pixels among each super-pixel are assigned to equal grey value; 2) based on the coarse-grain image, a saliency map is generated by merging primary features for each super-pixel. Thereby these super-pixel boundary discontinuities must be introduced into the final saliency map while they are merged into a larger region based on colour uniqueness and distribution.

Therefore, we need to explore more reliable saliency detection model which can avoid the mentioned drawbacks as many as possible. Itti et.al. proposed a top-down model, namely "center-surround", which uses center-surround differences between a "center" fine scale and a "surround" coarser scale to yield the saliency feature maps, which are formed by three elementary features: intensity, four broadly-tuned colour channel and LOI from oriented Gabor pyramids. Specially, we deeply analyse the single component, LOI and found it can provide profiles of salient regions as shown in Figure 2 (b). However, these profiles are consisted of some textures, in other words, these salient regions are not able to be homogeneously detected though they have been improved in terms of integrity. At least near the boundary of salient region, the loss of part has never ever occurred. Furthermore, if we can find a method, which can detect relatively homogeneous and uniform salient region instead of the above one, then LOI saliency map will

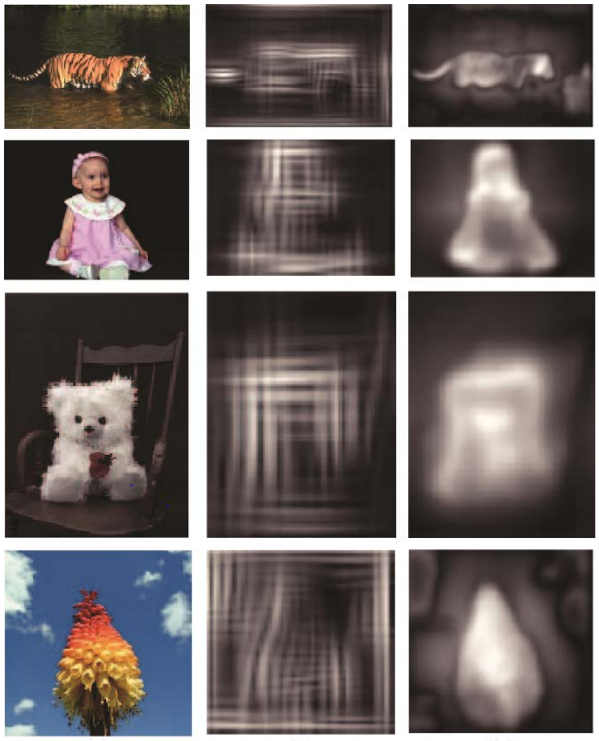

Fig. 2. *(a) source images; (b) saliency maps yielded by using Gabor filter on Gaussian pyramid sub-samples and bilinear interpolation; (c) saliency maps yielded by using reciprocal function filter on Gaussian pyramid sub-samples and bilinear interpolation.*

be very viable and reliable. A so-called reciprocal function defined by equation (4) can be used to solve this issue, which owns a property similar to the Gabor function as Figure 3, but outperforms the latter in highlighting the distinctive regions. As Figure 2 (c) shows, a remarkable improvement for the saliency maps can be found.

So far, we have found our visual attention model's idea, which is built based on three saliency contrasts: uniqueness, colour distribution and LOI. Here the two contrasts, uniqueness and colour distribution, are used for the proposed novel model to detect a coarse saliency map. Moreover, LOI contrast is used to supplement and tune the coarse saliency map. A smoothed and fine-grain saliency map can be attained in the end. Therefore, the procedure for our visual attention model includes Simple Linear Iterative Clustering (SLIC) super-pixel pre-segmentation [40], element uniqueness, element distribution, LOI to integrate uniqueness & distribution and LOI to compute the final saliency map. Our model is referred to as saliency filter with reciprocal function (SFR) since it is derived from the saliency filter model and reciprocal function filter, further it is formulated as follows.

3.1. Pre-segmentation

Similar to Perazzi's model [12], we first use SLIC super-pixels in Lab colour space to partition an image into spatial compact regions $S_i, i = 1,2,3,\cdots, L$ with relatively consistent size. By utilizing the *k*-means clustering approach, the super-pixel algorithm SLIC can efficiently generate compact and orderly super-pixels. An expected segmentation can be achieved while we only pass the expected number of pre-segmented super-pixels to the SLIC method. Therefore, for an input image, to observe more segmentation details we can easily use SLIC algorithm to yield its various segmentations at multi-scalessuper-pixelsuper-pixelsuper-pixel.

We compute colour uniqueness and distribution at each scale, and these super-pixels are used as elementary processing units for element uniqueness and distribute description. Let c_i *and* p_i denote the average colour and centre position of the *i*th super-pixel respectively, which are calculated by

$$c_i = \frac{\sum_{j=1}^{N_i} c_{ij}}{N_i}, p_i = \frac{\sum_{j=1}^{N_i} p_{ij}}{N_i} \quad (1)$$

where c_{ij}, p_{ij}, N_i are the *j*th colour, *j*th pixel and pixel number of the *i*th super-pixel respectively.

3.2 Colour contrast saliency

Each colour component belonging to a salient object should have a strong contrast to their surroundings [20]. Colour contrast saliency is yielded by the distinction of the *i*th super-pixel S_i with the centre position p_i and colour c_i compared to all other super-pixel S_j:

$$S_i^c = \sum_{j=1}^{N} \|c_i - c_j\|_2 w(p_i, p_j) \quad (2)$$

where $w(p_i,p_j)$ yields a local contrast term, which tries to emphasize that for two super-pixels belonging to the same object, there exists not only similar colours but also a nearer distance between them. This means that if there are similar colours and a nearer distance between S_i and S_j, the two super-pixels are probably merged into a larger region and super-pixel S_j provides contribution to super-pixel S_i. Therefore, the saliency of super-pixel S_j should be enhanced, whereas $w(p_i,p_j) = 1$ yields a global colour contrast, which cannot represent the sensitivity to local contrast variation. Generally, for the weight $w(p_i,p_j)$ in Eq.(2), Gaussian blurring kernel $w(p_i, p_j) = \frac{1}{\omega} exp\left(-\frac{1}{2\sigma_c^2}\|p_i - p_j\|_2\right)$can be used to approximate it, σ_c controls the range of colour contrast operator, which allows for a balance between local and global effects, i.e., given a smaller σ_c value, colour contrast saliency map yielded by equation (2) preserves more local regions or local effect. Conversely, for a larger σ_c value, global effect is enhanced in the yielded saliency map whereas local regions are suppressed. ω is the normalization factor ensuring $\sum_{j=1}^{N} w(p_i, p_j) = 1$.

3.3 Colour distribution

A super-pixel belonging to a salient object displays strong colour contrast to its surroundings, meanwhile, super-pixel belonging to the background also probably displays strong colour contrast to its surroundings [41]. However, from human visual perception colour super-pixels belonging to the salient object will be distributed within the more compact range in spatial structure and have a smaller spatial variance, whereas for this colour super-pixels belonging to the background will be distributed over the entire image and have a bigger spatial variance. We need to further define a

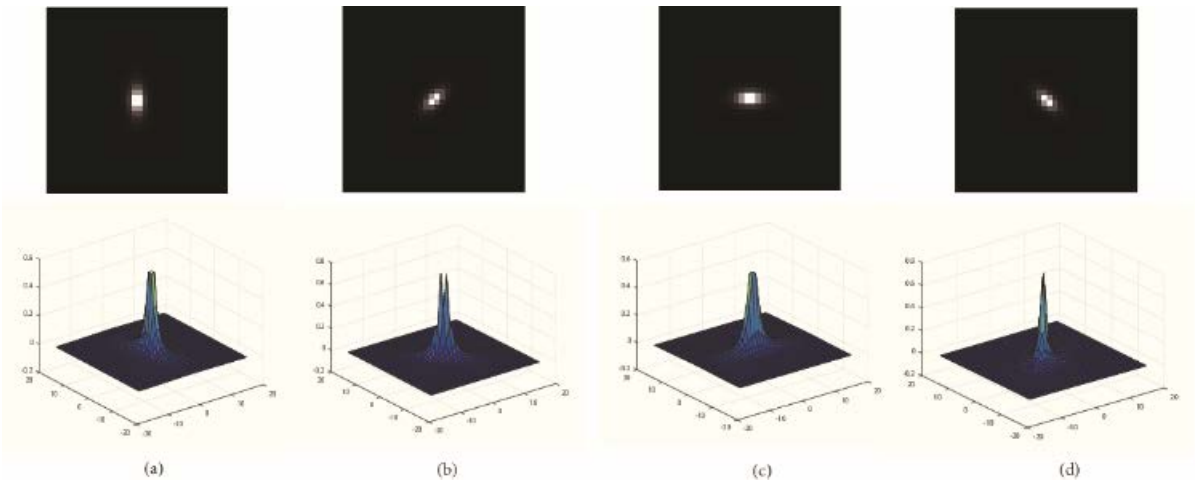

Fig. 3. *The 34×34 kernel masks yielded by the reciprocal function at the four orientations: 0,$\frac{\pi}{4}$,$\frac{\pi}{2}$,$\frac{3\pi}{4}$ are shown in columns (a), (b), (c), (d), respectively. The top row is their corresponding grey images, the bottom row is their corresponding surfaces.*

corresponding second measure to enhance the saliency of super-pixels belonging to the salient object and suppress the saliency of super-pixels belonging to the background.

For the *i*th super-pixel, we define its distribution measure by using the spatial variance D_i of its colour c_i, i.e., we measure its occurrence elsewhere in the image. As mentioned before, low variance indicates a spatially compact object, which should be considered more salient than spatially widely distributed elements. The spatial variance D_i is calculated by

$$D_i = \sum_{j=1}^{N} \left\| p_j - \mu_i \right\|_2 \omega_{ij}^{(c)} \tag{3}$$

where $\omega_{ij}^{(c)} = \frac{1}{W_i} exp\left(-\frac{1}{2\sigma_d^2}\left\| c_i - c_j \right\|_2\right)$ describes the similarity of colour c_i and colour c_j of super-pixels i and j, respectively, p_j is the position of super-pixel j, and $\mu_i = \sum_{j=1}^{N} \omega_{ij}^{(c)} p_j$ defines the weighted mean position of colour c_i, the parameter σ_d controls the colour sensitivity of the element distribution, a smaller σ_d suppresses colour effect between super-pixels so that distribution saliency map yielded by equation (3) is more homogeneous, conversely, the opposite effect is gained.

3.4 Orientation saliency

Itti's "centre-surround" model uses centre-surround differences between a "centre" fine scale and a "surround" coarser scale to yield the saliency feature maps, which are formed by three elementary features: intensity, four broadly-tuned colour channel and LOI from oriented Gabor pyramids. Contrasts as Figure 2 (b) are gained by using the LOI. For such LOI contrasts, these salient regions are roughly outlined, but there exist two main drawbacks: they are textured and not enough smooth near their boundaries. If these detected salient regions are homogeneous, uniform and enough smooth near the boundary areas, then we can use itti's model to correct lacks of parts near the regional boundary caused by saliency filter (SF) model. After investigating and analysing a large number of filters, a filter referred to as reciprocal function is used to substitute Gabor function in Itti's model and to attain the expected salient regions, which are homogeneous, uniform and smooth enough near their boundaries.

The reciprocal function filter is formulated by

$$\mathrm{f}(x,y) = \frac{1}{\sigma\left(\left(xcos(\theta)+ysin(\theta)\right)^2+\gamma^2\left(ycos(\theta)-xsin(\theta)\right)^2\right)+1} \tag{4}$$

where σ is control parameter, γ orientation curvature, and θ orientation angle. As shown in Figure 3, the four kernels are generated by the filter at the four orientations with the parameter values set above, which are similar to ones by the Gabor kernels.

Furthermore, a novel LOI-based detection model is developed. Let $I^{(i)}$ an intensity image with $I^{(i)} = (r + g + b)/3$, r, g, b being the red, green and blue channels of the source image I. $I^{(i)}$ is used to create a pyramid $I^{(i)}(m)$, where m is the scale. For each pyramid $I^{(i)}(m)$, a kernel of oriented reciprocal filter $\mathrm{f}(x, y, \theta)$ with size 34 × 34 is used to generate Gaussian pyramid $O(m, \theta)$, where $\theta \in \left\{0, \frac{\pi}{4}, \frac{\pi}{2}, \frac{3\pi}{4}\right\}$. Note that different from conventional Gaussian pyramid, $O(m, \theta)$ here is generated by the proposed reciprocal function filter $\mathrm{f}(x, y)$ substituting Gabor filter. Further, the local orientation contrast $O_c(c, s, \theta)$ at the orientation θ is calculated by

$$O_c(c,s,\theta) = |O(c,\theta)| \ominus |O(s,\theta)| \tag{5}$$

where $c \in \{3, 4\}$ and $s = c + \delta$, $\delta \in \{3,4\}$, $\ominus$ represents the summation operator pixel by pixel. Traversing c and s for equation (5), four intermediary maps are yielded at the orientation θ. Finally, we integrate the 16 maps at four orientations $0, \frac{\pi}{4}, \frac{\pi}{2}, \frac{3\pi}{4}$ into a single orientation saliency map by equation (6):

$$\bar{O} = \sum_{\theta \in \{0, \frac{\pi}{4}, \frac{\pi}{2}, \frac{3\pi}{4}\}} N\left(\oplus_{c=3}^{4} \oplus_{s=c+3}^{c+4} \left(N\left(O_c(c,s,\theta)\right)\right)\right) \tag{6}$$

where $N(\cdot)$ represents normalizing operator.

To better understand the process calculating single orientation saliency map, a block diagram is shown as Fig.4.

3.5. Saliency assignment

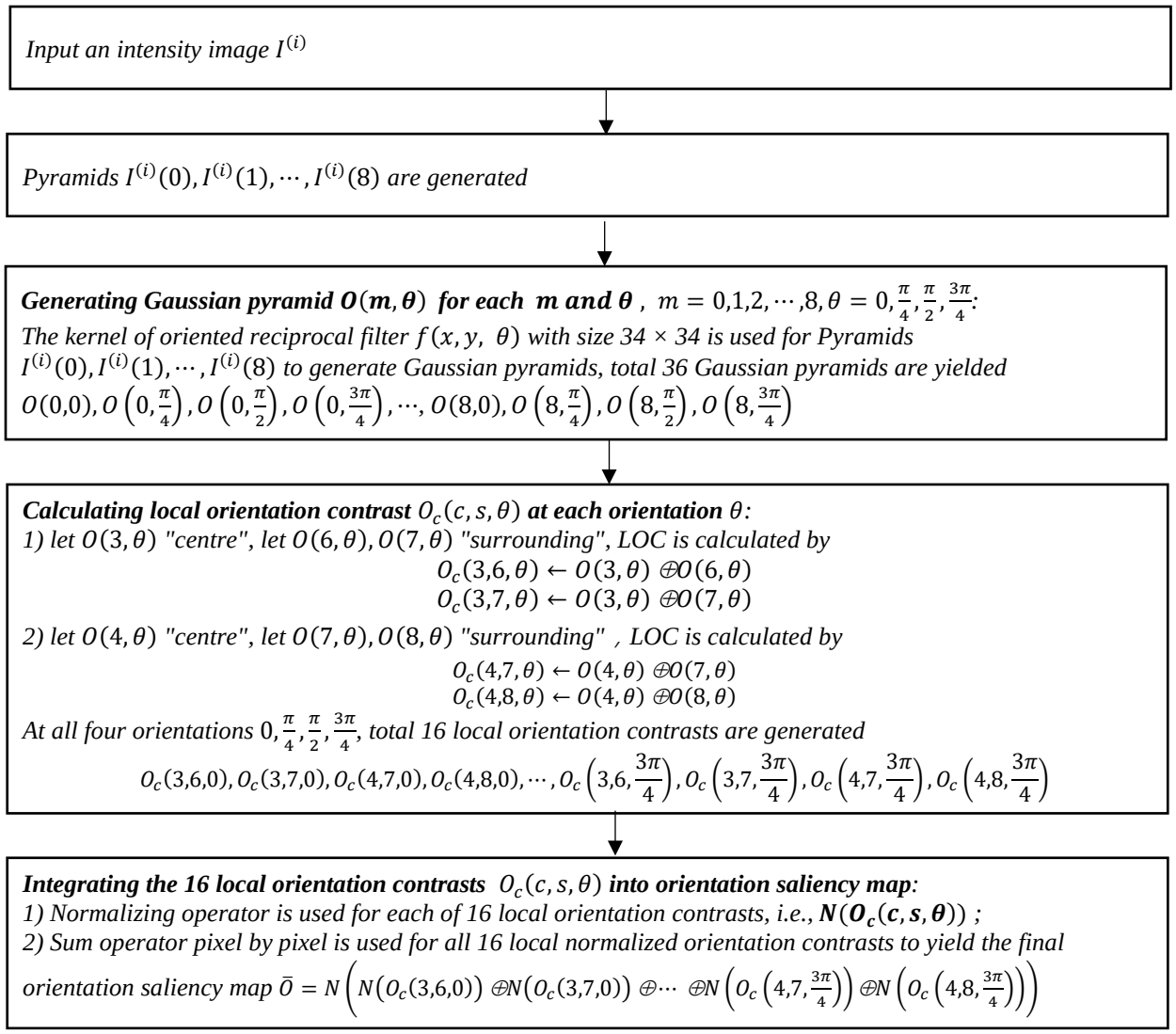

Fig.4. *The process calculating a single orientation saliency map*

Our framework generating the saliency map includes two steps: the first step is carried out to obtain the colour-distribution saliency map $S^{(cd)}$ by combining colour and distribution contrasts $S_i^{(c)}$ and $S_i^{(d)}$ for each super-pixel S_i; the second step is carried out to obtain the final saliency map S by combining $S^{(cd)}$ and LOI $\bar{O}$ pixel by pixel.

For the first step, similar to SF model by normalizing both $S_i^{(c)}$ and $S_i^{(d)}$ to the range [0, 1]. For each super-pixel S_i, its colour-distribution saliency map $S^{(cd)}$ is combined by

$$S_i^{(cd)} = S_i^{(c)} \odot \exp(-k \cdot D_i) \tag{7}$$

where $\odot$ is a multiply operator pixel by pixel, k is the scaling factor for the exponential. Traversing all N super-pixels, $S^{(cd)}$ is obtained.

For the second step, the final saliency map S is calculated by

$$S = S^{(cd)} \odot \bar{O} \tag{8}$$

Traversing all pixels in the source image I, S is obtained.

Figure 1 shows some experimental results on two benchmark datasets MSRK10 [7] and ECSSD [21], where images in (d) and (e) are the binary ones generated by using our method and the SF model with threshed $T = \frac{2 \times \sum_{i=1}^{W} \sum_{j=1}^{H} S(i,j)}{W \times H}$ respectively, images in (f) are the corresponding ground truths. Compared (d) and (e) to (f) respectively, the images in (e) are closer to (f) and gain remarkable improvements, especially near the boundary areas of salient regions our visual attention model demonstrates better integrity.

3.6. Saliency Map Fusion

For an input image, SLIC method can easily be applied and generate various super-pixel segmentations by specified the number of super-pixels. As Fig.5 shows, for a given image three different scale super-pixel segmentations are specified, where number of super-pixels are taken as 150, 250, 350 respectively. Further, the proposed model is used for the three super-pixel segmentations and generates three corresponding saliency maps, as Fig.5 (c) can be seen between which exist slight differences. These differences

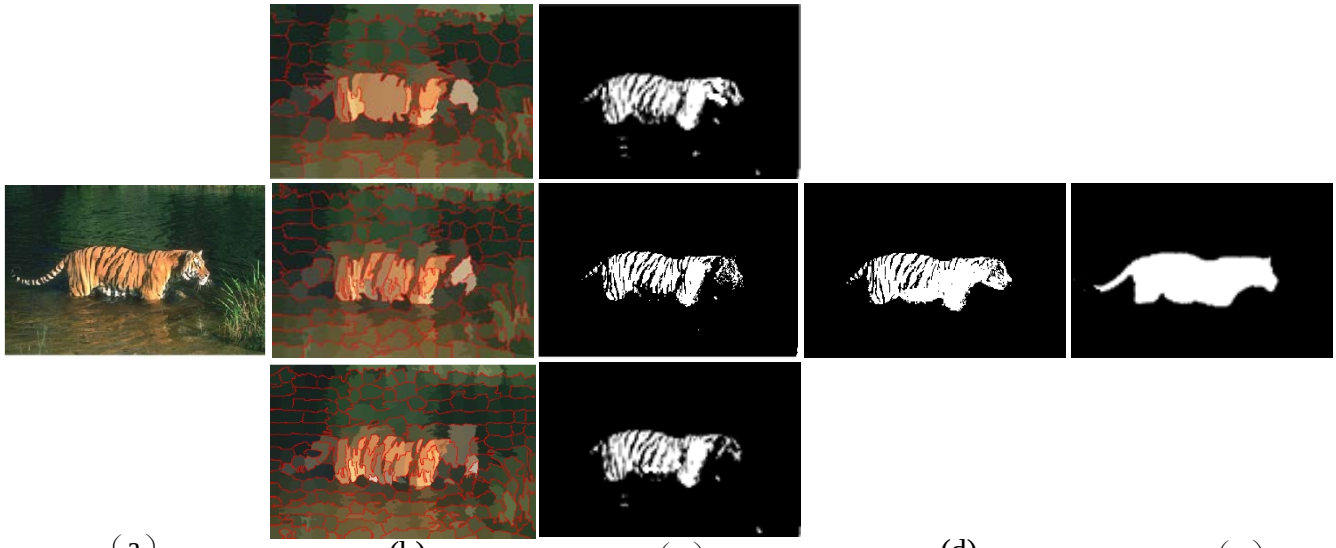

Fig.5. *Illustration of super-pixel segmentations and corresponding saliency maps for different scales. (a) source image; (b) super-pixel segmentations from top to bottom where number of super-pixels are taken as 150, 250,350 respectively; (c) saliency maps yielded for the three super-pixel segmentations; (d) saliency map optimized by fusing multiscale maps in cloumn (c); (e) Ground Truth*

demonstrate the diversity of saliency maps under multi-scale segmentations. This kind of diversity shows that there exists an opportunity to gain the optimum saliency map from a series of multiscale saliency maps. We proposed an optimizing strategy: 1) the proposed model is applied to each of N scale segmentations, and a series of N pre-refined saliency maps is generated, denoted as $\{S^{(1)},S^{(2)},\cdots,S^{(N)}\}$. 2) assuming that the final saliency map SA is a linear combination of the maps at each scale, then SA is formulated as follows:

$$SA=\sum_{k=1}^{N}\alpha_k S^{(k)}$$
$$s.t.\{\alpha_k\}_{k=1}^{N}=\arg\min_{\alpha_1,\alpha_2,\cdots,\alpha_N}\sum_{i\in I_v}\left\|S_i-\sum_k\alpha_k S_i^{(k)}\right\|_F^2. \quad (9)$$

Similar to [38], where the weights α_k can be learned by running a least-squares estimator over a validation dataset, indexed with I_v.

There are many options for saliency fusion. For example, a conditional random field (CRF) framework has been adopted in [27] to aggregate multiple saliency maps from different methods. Nevertheless, we have found that the linear combination of all saliency maps can already serve our purposes well and is capable of producing aggregated maps with a quality comparable to those obtained from more complicated techniques, as Fig.5 (d) shows the proposed saliency fusion strategy can not only generate smoother results but also well preserve salient object contours.

4. Experimental Results

4.1. Experimental Setup

4.1.1 Datasets: We evaluate the proposed algorithms on four benchmark datasets: MSRA10K [7], ECSSD [21], PASCAL-S [14] and DUT-OMRON [42]. MSRA10K contains 10,000 images with various objects. Most images contain only one salient object and the backgrounds are usually clear. The ECSSD dataset contains 1000 images with complex scenes and is considered much more challenging. DUT-OMRON contains 5,168 challenging images, each of which has one or more salient objects and relatively complex background. We have noticed that many saliency annotations in this dataset may be controversial among different human observers. As a result, none of the existing saliency models has a high accuracy on this dataset. Finally, we also evaluate models over PASCAL-S dataset, which was built using the validation set of the PASCAL VOC 2010 segmentation challenge. It contains 850 images with the ground truth labelled by 12 subjects.

4.1.2 Evaluation metrics: We use four universally-agreed standard metrics to evaluate our model: Precision-Recall (PR) curves, F-measure, the mean absolute error (MAE) and the area under ROC curve (AUC). For the sake of simplification, we use S to represent the predicted saliency map normalized to [0,255] and G to represent the ground-truth binary mask of salient objects. For a binary mask, we use $|\cdot|$ to represent the number of non-zero entries in the mask. The PR curve reflects the object retrieval performance in precision and recall by binarizing the final saliency map using different thresholds. For a saliency map S, we can convert it to a binary mask M and compute Precision and Recall by comparing M with ground-truth G:

$$precision=\frac{|\mathrm{M}\cap\mathrm{G}|}{|\mathrm{M}|},recall=\frac{|\mathrm{M}\cap\mathrm{G}|}{|\mathrm{G}|}. \quad (10)$$

Usually, neither Precision nor Recall can comprehensively evaluate the quality of a saliency map. To this end, the F-measure is proposed as a weighted harmonic mean of them with a non-negative weight β:

$$F_\beta=\frac{(1+\beta^2)precision\times recall}{\beta^2 precision+recall} \quad (11)$$

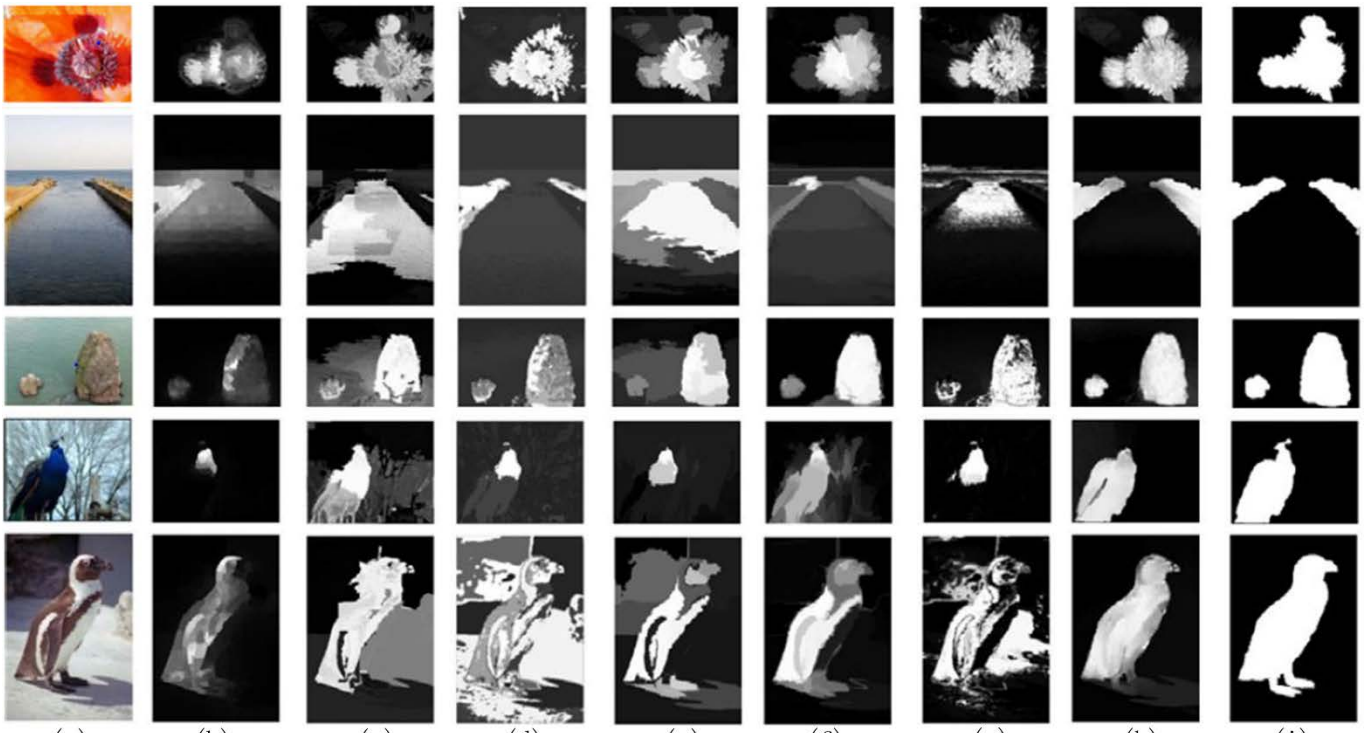

Fig. 6. *Visual comparison of saliency maps generated from 7 different methods, including ours. From left to right (columns): input, FT [44], RC [7], GC [46], HS [21], DRFI [43], LEGS [45], the proposed method and the ground truth.*

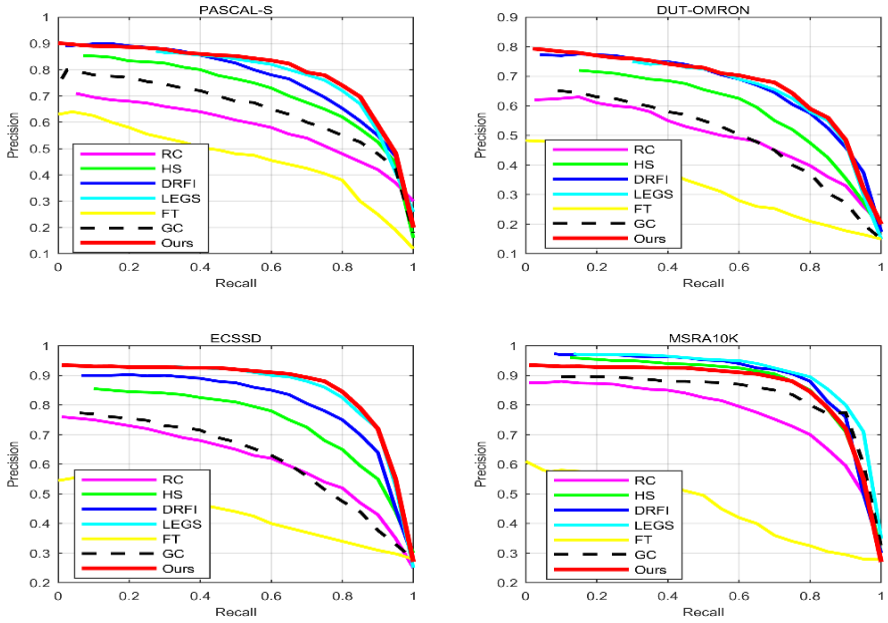

Fig. 7. *Precision-recall curves of different saliency detection methods on 4 benchmark datasets. Overall, the proposed approach performs well with higher precision in the case of a fixed recall.*

As suggested by many salient object detection works (e.g., [7], [8], [12]), β^2 is set to 0.3 to raise more importance to the Precision value. The AUC value is the percentage of the area under the ROC curve, which indicates how well the saliency map predicts the real salient objects. While ROC is a two-dimensional representation of performance for a model, the AUC distils this information into a single scalar. As the name implies, it is calculated as the area under the ROC curve. A perfect model will score an AUC of 1, while random guessing will score an AUC around 0.5. Let $\bar{S}$ and $\bar{G}$ denote the continuous saliency map and the ground truth that are normalized to [0, 1]. The MAE score can be computed by

$$MAE = \frac{1}{W \times H} \sum_{x=1}^{W} \sum_{y=1}^{H} |\bar{S}(x,y) - \bar{G}(x,y)| \qquad (12)$$

As stated in [21], this metric favours method that successfully detects salient pixels but fails to detect non-

salient regions over method that successfully detects non-salient pixels but makes mistakes in determining the salient ones.

Table 1 *Quantitative comparison for ECSSD, DUT-OMRON, PASCAL-S and MSRA10K datasets.*

Dataset	Metric	Ours	HS	DRFI	FT	GC	RC	LEGS
ECSSD	AUC	0.9003	0.8293	0.877	0.6443	0.7655	0.833	—
	MAE	0.1601	0.2064	0.1841	0.2859	0.2382	0.187	0.191
DUT-OMRON	AUC	0.9432	0.8602	0.9335	0.682	0.7956	0.8592	—
	MAE	0.0845	0.1568	0.0978	0.1761	0.1675	0.29	—
PASCAL-S	AUC	0.9154	0.8267	0.881	0.6181	0.7995	0.8379	—
	MAE	0.1645	0.2376	0.2351	0.3297	0.2655	0.225	0.17
MSRA10K	AUC	0.9477	0.8108	0.8624	0.6004	0.7178	0.8238	—
	MAE	0.1503	0.2297	0.2163	0.2835	0.2523	0.242	—

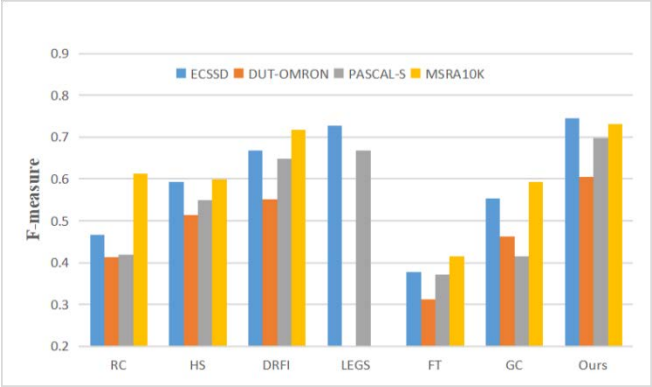

Fig. 8. *Average F-measure for the compared models on all four benchmark datasets.*

4.1.3 *Implementation:* All the experiments are carried out using MATLAB on a desktop computer with an Intel i7-3770 CPU (3.4GHz) and 32GB RAM. We empirically set σ_c = 0.25 (in Eq.2) and σ_d = 20 (in Eq. 3). In the orientation saliency model, we set parameters in Eq.4 as σ = 2.33, γ = 1, $\theta \epsilon \left\{0, \frac{\pi}{4}, \frac{\pi}{2}, \frac{3\pi}{4}\right\}$. The Gaussian pyramid $O(m, \theta)$, where $m \in \{0, \cdots, 8\}$, is gained by the kernel of oriented reciprocal filter $f(x, y, \theta)$ with size 34 × 34 in all our experiments for the given intensity image $I^{(i)}(m)$. According to [12], we use k=6 in Eq. 7. For the saliency map fusion, here we take number of super-pixels under three scales of segmentations as 150, 250, 350 respectively.

4.2. Comparison to other methods

A qualitative as well as a quantitative evaluation is done in order to measure the performance of the proposed model, and is compared with the existing approaches. For fair comparison, we use the original source codes or the provided saliency detection results in the literature.

We compared our model with several state-of-the-art models, including RC [7], HS [21], DRFI [43], FT [44], LEGS [45] and GC [46]. Among these methods, LEGS (deep learning) and DRFI (random forest regressor) are learning-based methods; RC and HS are based on global contrast; the classic methods FT are included as a baseline. Most of the saliency maps associated with the competing approaches can be obtained by running their publicly availabel source code using the default experimental configurations.

A visual comparison is given in Fig. 6. As can be seen, our method performs well in a variety of challenging cases, in which the boundary of salient region can be continuously reserved.

As part of the quantitative evaluation, we first use the Precision and Recall (P-R) curve to evaluate all the methods. We set the fixed threshold from 0 to 255 with an increment of 5 for a saliency map with consistent grey value, thus producing 52 binary masks. Using the pixel-wise ground truth data, 52 pairs of average P-R values of all the images included in the test datasets are computed. Fig. 7 shows the P-R curves where several state-of-the-art methods and the proposed algorithms perform well.

Fig. 9. *PR curves of the proposed model with the reciprocal function filter, Gabor filter and without multiscale fusion, respectively.*

We show the average F_β values of all the competing approaches on the four benchmark datasets in Fig. 8. From Fig. 8, we observe that the proposed method achieves a better performance than the other ones in most cases.

Most specifically, Table 1 reports their quantitative saliency detection performance w.r.t. the evaluation metrics (i.e., AUC and MAE) on the four benchmark datasets. From Figure 6 and Table 1, it is clearly seen that our approach performs favourably against the state-of-the-art methods in most cases.

4.3. Analysis of Proposed Approach

4.2.1 Effectiveness of the reciprocal function: As discussed in Sec. 3.4, the reciprocal function filter we proposed to substitute Gabor funstion in Itti's model can detect the expected salient regions, with homogeneous, uniform and smooth enough boundaries. To validate its effectiveness, we have evaluated the performance of our final saliency map with Gabor filter and with reciprocal function filter using the testing images in the PASCAL-S dataset. As shown in Fig.9, the model using reciprocal function performs much better than the one using Gabor filter in terms of the PR curve.

*4.2.2 Effectiveness of multiscale decomposition: O*ur method exploits information from multiple scales of image segmentation. The results are also shown in Fig.9. The performance of a single segmentation scale is not comparable to the performance of the fused model. The aggregated saliency map improves the average precision and the recall rate when it is compared with the result from the best-performing single scale.

5. Conclusion

In this paper, we propose a novel model for salient object detection in which saliency map is computed as a combination of the colour and its distribution-based saliency and orientation saliency. The proposed method is based on super-pixel segmentation to map the regional feature vector to a saliency score. We combine colour uniqueness and distribution with local orientation information and define a reciprocal function to substitute the Gabor function used in LOI, and have proved that the reciprocal function could produce a remarkable improvement for the saliency detection. Saliency scores across multiple layers are finally fused to produce the saliency map. We evaluate the proposed method extensively on the four benchmark datasets and make comparison with 6 state-of-the-art algorithms. Experimental results verify the detection accuracy and efficiency of our method.

In the future, we shall continue to survey all types of detection and extraction methods based on salient object and further explore the new methods and their application in practice.

Acknowledgements

This work was partly supported by the National Natural Science Foundation of China (Grant No. 61672291) and the Beijige Open Fund of Jiangsu Meteorology Science Research Institute (Grant No. BJG201504). Also, the first author would like to acknowledge the support provided by the EU Erasmus Mundus project SmartLink (EACEA, 2014-0858) to carry out this research at the University of the West of Scotland, UK.

References

[1] Donoser, M., Urschler, M., Hirzer, M., Bischof, H.: 'Saliency driven total variation segmentation'. In ICCV, 2009.

[2] Itti, L.: Automatic foveation for video compression using a neurobiological model of visual attention. IEEE TIP, 2004.

[3] Kanan, C., Cottrell, G. W.: Robust classification of objects, faces, and flowers using natural image statistics, in CVPR, 2010, pp. 24722479.

[4] Chen, T., Cheng, M.-M., Tan, P., Shamir, A., Hu, S.M.: Sketch2Photo: internet image montage [J], Acm Transactions on Graphics, 2009, 28(5):1-10.

[5] Parkhurst, D., Law, K. and Niebur, E.: Modeling the role of salience in the allocation of overt visual attention. *Vision research*, 42(1):107–123, 2002.

[6] Itti, L. and Koch, C.: Computational modelling of visual attention. Nature reviews neuroscience, vol. 2, no. 3, pp. 194–203, 2001.

[7] Cheng, M.M., Zhang, G.X., Mitra, N.J.: Global contrast based salient region detection, IEEE Conference on Computer Vision and Pattern Recognition, 37(3) (2011) 569 - 582.

[8] Liu, T., Yuan, Z., Sun, J., Wang, J., Zheng, N., Tang, X., Shum, H.- Y.: Learning to detect a salient object, IEEE TPAMI, vol. 33, no. 2, pp. 353367, 2011.

[9] Yang, C., Zhang, L., Lu, H., Ruan, X. and Yang, M.-H.: Saliency detection via graph-based manifold ranking, in Proc. IEEE Conf. CVPR, 2013, pp. 3166–3173.

[10] Han, J., Zhang, D., Hu, X., Guo, L., Ren, J., Wu, F.: Background Prior-Based Salient Object Detection via Deep Reconstruction Residual. IEEE Trans, CSVT, 25(8): 1309-1321, 2015.

[11] Wei, Y., Wen, F., Zhu, W., Sun, J.: Geodesic saliency using background priors, ECCV, 2012.

[12] Perazzi, F., Krahenbuhl, P., Pritch, Y., Hornung, A .: Saliency filters: Contrast based filtering for salient region detection, CVPR, 2012.

[13] Chang, K.-Y., Liu, T.-L., Chen, H.-T., Lai, S.-H.: Fusing generic objectness and visual saliency for salient object detection, ICCV, 2011.

[14] Li, Y., Hou, X., Koch, C., Rehg, J. M., Yuille, A. L.: The secrets of salient object segmentation, CVPR, 2014.

[15] LeCun, Y., Bottou, L., Bengio, Y. and Haffner, P.: Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11): 2278-2324, 1998.

[16] He, K., Zhang, X., Ren, S. and Sun, J.: Spatial pyramid pooling in deep convolutional networks for visual recognition. in *ECCV*, 2014.

[17] Girshick, R., Donahue, J., Darrell, T. and Malik, J.: Rich feature hierarchies for accurate object detection and semantic segmentation. in *CVPR*, 2014.

[18] Itti, L., Koch, C., Niebur, E .: A model of saliency-based visual attention for rapid scene analysis, PAMI, 20(11): 1254-1259, 1998.

[19] Borji, A., Sihite, D. N., Itti, L.: Salient Object Detection: A Benchmark, IEEE Transactions on Image Processing 24(12)(2015) 414-429.

[20] Borji, A., Cheng, M. M., Jiang, H., et al.: Salient Object Detection: A Survey, Eprint Arxiv 16(7) (2014)3118.

[21] Yan, Q., Xu, L., Shi, J., Jia, J.: Hierarchical saliency detection, Computer Vision and Pattern Recognition IEEE, 2013:1155-1162.

[22] Jiang, B., Zhang, L., Lu, H., et al.: Saliency Detection via Absorbing Markov Chain, IEEE International Conference on Computer Vision, IEEE Computer Society, 2013:1665-1672.

[23] Li, X., Lu, H., Zhang, L., et al.: Saliency Detection via Dense and Sparse Reconstruction, IEEE International Conference on Computer Vision. IEEE, 2013:2976-2983.

[24] Zhu, W., Liang, S., Wei, Y., et al.: Saliency Optimization from Robust Background Detection, IEEE Conference on Computer Vision and Pattern Recognition. IEEE Computer Society, 2014:2814-2821.

[25] Liu, T., Yuan, Z., Sun, J., Wang, J., Zheng, N., Tang, X., Shum, H.- Y.: Learning to detect a salient object, IEEE TPAMI, vol. 33, no. 2, pp. 353367, 2011.

[26] Khuwuthyakorn, P., Robles-Kelly, A., Zhou, J.: Object of interest detection by saliency learning, ECCV, 2010.

[27] Lu, S., Mahadevan, V., Vasconcelos, N.: Learning optimal seeds for diffusion-based salient object detection, CVPR, 2014.

[28] Kim, J., Han, D., Tai, Y.-W., Kim, J.: Salient region detection via high-dimensional colour transform, CVPR, 2014.

[29] Mehrani, P., Veksler, O.: Saliency segmentation based on learning and graph cut refinement, BMVC, 2010.

[30] Li, Y., Hou, X., Koch, C., Rehg, J. M., Yuille, A. L.: The secrets of salient object segmentation, CVPR, 2014.

[31] Carreira, J., Sminchisescu, C.: Constrained parametric min-cuts for automatic object segmentation, CVPR, 2010, pp. 32413248.

[32] Wang, P., Wang, J., Zeng, G., Feng, J., Zha, H., Li, S.: Salient object detection for searched web images via global saliency, CVPR, 2012, pp. 31943201.
[33] Moosmann, F., Larlus, D., Jurie, F.: Learning saliency maps for object categorization, EECVW, 2006.
[34] He S, Lau R W, Liu W, et al. SuperCNN: A Superpixelwise Convolutional Neural Network for Salient Object Detection [J]. International Journal of Computer Vision, 2015, 115(3):330-344.
[35] Huang F, Qi J, Lu H, et al. Salient object detection via multiple instance learning[J]. IEEE Transactions on Image Processing, 2017, 26(4): 1911-1922.
[36] Jiang H, Cheng M M, Li S J, et al. Joint Salient Object Detection and Existence Prediction[J]. Front. Comput. Sci, 2017.
[37] Li X, Zhao L, Wei L, et al. DeepSaliency: Multi-Task Deep Neural Network Model for Salient Object Detection [J]. IEEE Transactions on Image Processing, 2016, 25(8):3919-3930.
[38] Li G, Yu Y. Deep Contrast Learning for Salient Object Detection [J]. 2016:478-487.
[39] Liu N, Han J. DHSNet: Deep Hierarchical Saliency Network for Salient Object Detection[C]// Computer Vision and Pattern Recognition. IEEE, 2016:678-686.
[40] Achanta, R., Shaji, A., Smith, K. et al.: SLIC superpixels compared to state-of-the-art superpixel methods, IEEE Transactions on Pattern Analysis and Machine Intelligence, 2012, 34(11):2274.
[41] Collins, R.T.: Mean-shift blob tracking through scale space, in: IEEE Conference on Computer Vision and Pattern Recognition, vol. 2, 2003, pp. 234-240.
[42] Yang, C., Zhang, L., Lu, H., et al.: Saliency detection via graph-based manifold ranking. In Computer Vision and Pattern Recognition (CVPR), 2013 IEEE Conference on, pages 3166–3173. IEEE, 2013.
[43] Jiang, P., Ling, H., Yu, J., Peng, J.: Salient region detection by ufo: Uniqueness, focusness and objectness, Computer Vision (ICCV), 2013 IEEE International Conference on, pages 19761983. IEEE, 2013.
[44] Achanta, R., Hemami, S., Estrada, F., et al.: Frequency-tuned salient region detection[C]// Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on. IEEE, 2009:1597-1604.
[45] Wang L., Lu, H., Ruan, X., Yang, M.-H.: Deep networks for saliency detection via local estimation and global search. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 31833192, 2015.
[46] Cheng, M. M., Warrell, J., Lin, W.Y., et al.: Efficient Salient Region Detection with Soft Image Abstraction[C]// IEEE International Conference on Computer Vision. IEEE Computer Society, 2013:1529-1536.
[47] Hou Q, Cheng M M, Hu X, et al. Deeply supervised salient object detection with short connections[C]//Computer Vision and Pattern Recognition (CVPR), 2017 IEEE Conference on. IEEE, 2017: 5300-5309.